On MMSE Properties and I-MMSE Implications in Parallel MIMO Gaussian Channels

Ronit Bustin
Dept. Electrical Engineering
Technion−IIT
Technion City, Haifa 32000
Israel
Email: bustin@tx.technion.ac.il

Miquel Payaró
Dept. of Radiocommunications
Centre Tecnològic de Telecomunicacions
de Catalunya (CTTC)
Castelldefels, Barcelona, Spain
Email: miquel.payaro@cttc.es

Daniel P. Palomar
Dept. of Elec. and Comp. Engineering
Hong Kong University of Science
and Technology
Clear Water Bay, Kowloon, Hong Kong
Email: palomar@ust.hk

Shlomo Shamai (Shitz)
Dept. Electrical Engineering
Technion−IIT
Technion City, Haifa 32000
Israel
Email: sshlomo@ee.technion.ac.il

***Abstract*—This paper extends the "single crossing point" property of the scalar MMSE function, derived by Guo, Shamai and Verdú (first presented in ISIT 2008), to the parallel *degraded* MIMO scenario. It is shown that the matrix $\boldsymbol{Q}(t)$, which is the difference between the MMSE assuming a Gaussian input and the MMSE assuming an arbitrary input, has, at most, a single crossing point for each of its eigenvalues. Together with the I-MMSE relationship, a fundamental connection between Information Theory and Estimation Theory, this new property is employed to derive results in Information Theory. As a simple application of this property we provide an alternative converse proof for the broadcast channel (BC) capacity region under covariance constraint in this specific setting.**

I. Introduction

A fundamental relationship between estimation theory and information theory for Gaussian channels was presented in [1]; in particular, it was shown that for the MIMO standard Gaussian channel,

$$\boldsymbol{Y} = \sqrt{\text{snr}}\boldsymbol{H}\boldsymbol{X} + \boldsymbol{N} \tag{1}$$

where $\boldsymbol{N}$ is a standard Gaussian n-dimensional random vector and $\boldsymbol{H}$ is a fixed channel matrix known to the receiver, then regardless of the input distribution on $\boldsymbol{X}$, the mutual information and the minimum mean-square error (MMSE) are related (assuming real-valued inputs/outputs) by

$$\frac{d}{d\text{snr}} I(\boldsymbol{X}; \sqrt{\text{snr}}\boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}) = \frac{1}{2}\mathbb{E}\{\| \boldsymbol{H}\boldsymbol{X} - \boldsymbol{H}\mathbb{E}\{\boldsymbol{X}|\sqrt{\text{snr}}\boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}\} \|^2\}.$$

Here $\mathbb{E}\{X|Y\}$ stands for the conditional mean of X given Y. This fundamental relationship and its generalizations [1], [2], referred to as the I-MMSE relationships, have already been shown to be useful in several aspects of information theory: providing insightful proofs for entropy power inequalities [3], revealing the mercury/waterfilling optimal power allocation over a set of parallel Gaussian channels [4] and recently generalizing this result to MIMO Gaussian channels in [5], tackling the weighted sum-MSE maximization in MIMO broadcast channels [6], illuminating extrinsic information of good codes [7], and enabling a simple proof of the monotonicity of the non-Gaussianness of independent random variables [8]. In [9], [10] and later in [11] it has been shown that using this relationship one can provide insightful and simple proofs for multi-user single antenna problems such as the BC, the secrecy capacity problem, and the multi-receiver secrecy capacity region. In [12] this approach has been extended to the MIMO Gaussian wiretap channel yielding a closed form expression for the secrecy capacity. In order to provide the converse proof of the BC capacity region in [9], [10], the authors proved an inherent property of the MMSE, the "single crossing point" property: as a function of *snr*, the MMSE of the Gaussian input distribution and the MMSE of an arbitrary input distribution intersect at most once. This property is stronger than required in order to prove the BC capacity region, however it is an interesting property on its own.

Motivated by this approach, our goal is to examine the properties of the MMSE matrix in the MIMO scenario, and relate these properties to the mutual information. As an initial model we have chosen the following simplified parallel channel,

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N} \tag{2}$$

where $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{N}$ are *n*-dimensional random vectors, and $\boldsymbol{N}$ is standard Gaussian. $\boldsymbol{H}$ is assumed diagonal and positive semidefinite. Note that $\boldsymbol{X}$ is not necessarily composed of independent components, in which case the "single crossing point" property can be deduced from the "single crossing point" property of the individual components.

As pointed out earlier, in the scalar channel scenario we have seen that a "single crossing point" between the MMSE of the Gaussian input and the MMSE of an arbitrary input distribution exists as a function of *snr*. The current more complex scenario, in which we have diagonal channel matrices, raises two questions: What scalar function of the MMSE matrix should we examine for an analogous property to the "single crossing point"? And, along what $n \times n$-dimensional path should we look (since in our parallel MIMO scenario there are multiple $n \times n$-dimensional paths between every two channel

The work of R. Bustin, M. Payaró and S. Shamai has been supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++. The work of R. Bustin and S. Shamai was also supported by the Israel Science Foundation.

The work of M. Payaró was also partially supported by the Spanish Government under project TEC2008-06327-C03-03/TEC (FBMC-SILAN).

The work of D. P. Palomar has been supported in part by the Hong Kong Research Grant DAG_S08/09.EG05.

matrices)? For consistency, even before answering these two questions, we would like to emphasize that the path will be parameterized through the scalar parameter t, thus instead of $\boldsymbol{H}$ we will write $\boldsymbol{R}(t)$, to avoid confusion.

The paper is organized as follows: section II contains the most basic definitions used in this work. Section III details our choice of path and gives some preliminary results used to prove our primary result. Section IV contains our main result, which is an extension of the single crossing point property to the parallel *degraded* MIMO scenario. Section V connects the result of the previous section to the mutual information using the I-MMSE relationship. Finally, section VI demonstrates how we can use this property to provide an alternative converse proof for the BC capacity region under covariance constraint in the parallel *degraded* MIMO setting.

II. Definitions

We now formally give the definition of the MMSE matrix:

$$\begin{aligned}\boldsymbol{E}(t) &= \mathbb{E}\{(\boldsymbol{X}-\mathbb{E}\{\boldsymbol{X}|\boldsymbol{R}(t)\boldsymbol{X}+\boldsymbol{N}\}) \\ &\quad (\boldsymbol{X}-\mathbb{E}\{\boldsymbol{X}|\boldsymbol{R}(t)\boldsymbol{X}+\boldsymbol{N}\})^T\} \end{aligned} \tag{3}$$

where $\boldsymbol{R}(t)$ corresponds to the channel matrix $\boldsymbol{H}$. The parameter t determines the channel matrix, thus the new variable $\boldsymbol{R}(t)$ comes to highlight the dependence of the channel on the parameter t.

In this work we use the following I-MMSE relationship derived by Palomar and Verdú in [2]:

$$\nabla_{\boldsymbol{H}} I(\boldsymbol{X};\boldsymbol{H}\boldsymbol{X}+\boldsymbol{N}) = \boldsymbol{H}\boldsymbol{E} \tag{4}$$

where $\boldsymbol{H}$ is a fixed known channel matrix, and $\boldsymbol{N}$ is a standard Gaussian additive noise. Rewriting this relationship as an integration along a path $\boldsymbol{R}(t)$ from $t=0$ to t' results with the following expression:

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}(t')) &= I(\boldsymbol{X};\boldsymbol{R}(t')\boldsymbol{X}+\boldsymbol{N}) \\ &= \int_{t=0}^{t'} \mathrm{tr}\left(\left(\boldsymbol{R}(t)\boldsymbol{E}(t)\right)^T \boldsymbol{R}'(t)\right) dt \end{aligned} \tag{5}$$

where $\boldsymbol{R}'(t) = \frac{d\boldsymbol{R}(t)}{dt}$.

In this work we specifically examine the properties of the difference between the MMSE resulting from an arbitrary input distribution and a Gaussian input distribution (not necessarily having the same covariance matrix). As such, we require the following definition:

$$\boldsymbol{Q}(t) = \boldsymbol{E}_G(t) - \boldsymbol{E}(t) \tag{6}$$

where we have denoted $\boldsymbol{E}_G(t)$ the MMSE matrix assuming a general Gaussian input distribution.

III. Preliminaries

We begin this section by presenting our choice of path, that is, we provide an answer to the second question presented in the Introduction: Along what $n \times n$-dimensional path should we look? In the scalar "single crossing point" property [9], [10], the MMSE is given as a function of *snr*, a non-negative value. The change in MMSE is examined as *snr* monotonically increases. When switching to the MIMO scenario, our choice was to mimic the properties of the scalar scenario, that is, we show that there exists a non-negative, monotonically non-decreasing path between any two diagonal matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, such that $\boldsymbol{0} \preceq \boldsymbol{H}_1 \preceq \boldsymbol{H}_2$, as given in the following lemma.

Lemma 1. *For any two diagonal matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, such that $\boldsymbol{0} \preceq \boldsymbol{H}_1 \preceq \boldsymbol{H}_2$, there exists a non-negative, monotonically non-decreasing path $\boldsymbol{R}(t)$ for all $t \in [0,1]$ such that the following holds:*

$$\begin{aligned} \boldsymbol{R}(t=0) &= \boldsymbol{0} \\ \boldsymbol{R}(t_1) &= \boldsymbol{H}_1 \\ \textit{and}\quad \boldsymbol{R}(t_2=1) &= \boldsymbol{H}_2 \end{aligned} \tag{7}$$

where $0 < t_1 < t_2 = 1$.

Proof: We need to define a function, $g_i(t)$, for each diagonal element i. It suffices to choose any non-negative function $f_i(t)$ such that the area from 0 to t_1 will equal $[\boldsymbol{H}_1]_{ii}$ and the area from t_1 to t_2 will equal $[\boldsymbol{H}_2]_{ii} - [\boldsymbol{H}_1]_{ii}$. Given that, we can set the function to be $g_i(t) = \int_0^t f_i(t')dt'$. The entire path, $\boldsymbol{R}(t)$, will be given by:

$$\boldsymbol{R}(t) = \mathrm{diag}\{g_1(t), ..., g_n(t)\}. \tag{8}$$

As required, this path passes between the zero matrix at $t=0$, $\boldsymbol{H}_1$ at t_1 and $\boldsymbol{H}_2$ at $t_2 = 1$. Since $f_i(t)$ are chosen non-negative for all i we have a non-negative and monotonically non-decreasing path for all $t \in [0,1]$. ∎

We now turn to provide some preliminary results that will be shown central in the sequel.

Lemma 2 ([13, Ch. 4, Sec. 11])**.** *Let λ_i and $\mathbf{u}_i$ indicate the i-th eigenvalue (assumed of multiplicity 1) of the matrix $\boldsymbol{Z}$ and its corresponding eigenvector, respectively. Then, it follows that*

$$\mathsf{D}_{\boldsymbol{Z}}\lambda_i = \mathbf{u}_i^T \otimes \mathbf{u}_i^T \tag{9}$$

where D is the Jacobian operator, whose definition can be found in [13].

Corollary 1. *If the matrix $\boldsymbol{Z}$ depends on a real scalar parameter τ,* i.e., *$\boldsymbol{Z} = \boldsymbol{Z}(\tau)$, then, assuming λ_i is an eigenvalue of multiplicity 1, applying the chain rule, we get:*

$$\frac{\mathrm{d}\lambda_i}{\mathrm{d}\tau} = \mathsf{D}_{\boldsymbol{Z}(\tau)}\lambda_i \mathsf{D}_\tau \boldsymbol{Z}(\tau) \tag{10}$$

$$= \left(\mathbf{u}_i^T \otimes \mathbf{u}_i^T\right)\mathsf{vec}\left(\boldsymbol{Z}'(\tau)\right) \tag{11}$$

$$= \mathbf{u}_i^T \boldsymbol{Z}'(\tau)\mathbf{u}_i \tag{12}$$

where the last equality follows from [13, Ch. 2, Th. 2.2].

Corollary 2. *If, given a $\tau = \tau_0$, the matrix $\boldsymbol{Z}(\tau_0)$ is diagonal, we can always take $[\mathbf{u}_i]_j = \delta_{ij}$ and, thus, the result in Corollary 1 particularizes to*

$$\left.\frac{\mathrm{d}\lambda_i(\boldsymbol{Z}(\tau))}{\mathrm{d}\tau}\right|_{\tau=\tau_0} = \left.\frac{\mathrm{d}[\boldsymbol{Z}(\tau)]_{ii}}{\mathrm{d}\tau}\right|_{\tau=\tau_0}. \tag{13}$$

Remark 1. *Observe that the results in Lemma 2 and Corollary 1 are valid only for the case where the multiplicity of λ_i is equal to 1. However, as explained in [13, Ch. 8, Sec. 12] and formally stated in [13, Ch. 8, Sec. 12, Th. 13], in our case, the result in Corollary 2 can be applied directly to the case where the multiplicity of λ_i is greater than 1.*

IV. Single Crossing Point for each Eigenvalue

Before stating our primary result we require a lower bound on the matrix $\boldsymbol{Q}(t)$, defined in (6), which is given in the following lemma.

Lemma 3. *The following lower bound holds:*

$$\boldsymbol{Q}'(t) \succeq 2\left(\boldsymbol{E}(t)\boldsymbol{B}(t)\boldsymbol{E}^T(t) - \boldsymbol{E}_G(t)\boldsymbol{B}(t)\boldsymbol{E}_G^T(t)\right) \quad (14)$$

where $\boldsymbol{B}(t) = \boldsymbol{R}(t)\boldsymbol{R}'(t)$ is a diagonal matrix.

Proof: We first provide the derivative of the MMSE with respect to the parameter t. Using the chain rule given in [14, equations (65-66)],

$$\begin{aligned} D_t\boldsymbol{E}_{ij}(t) &= D_H\boldsymbol{E}_{ij}(t)D_t\boldsymbol{R}(t) \\ &= \text{tr}\left(\frac{\partial \boldsymbol{E}_{ij}(t)}{\partial \boldsymbol{R}(t)}^T \boldsymbol{R}'(t).\right) \end{aligned} \quad (15)$$

Since $\boldsymbol{R}(t)$ is diagonal, the last expression can be further simplified to

$$D_t\boldsymbol{E}_{ij}(t) = \sum_l \frac{\partial \boldsymbol{E}_{ij}(t)}{\partial \boldsymbol{R}_{ll}(t)}\boldsymbol{R}'_{ll}(t). \quad (16)$$

Using the result ([14, eq. (131)]),

$$\begin{aligned} D_{\boldsymbol{R}_{ll}(t)}\boldsymbol{E}_{ij}(t) &= -\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_{jl}\left[\boldsymbol{\phi}_X(\boldsymbol{Y})\boldsymbol{R}(t)^T\right]_{il} \\ &\quad +\boldsymbol{\phi}_X(\boldsymbol{Y})_{il}\left[\boldsymbol{\phi}_X(\boldsymbol{Y})\boldsymbol{R}(t)^T\right]_{jl}\} \\ &= -\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_{jl}\boldsymbol{\phi}_X(\boldsymbol{Y})_{il}\boldsymbol{R}(t)_{ll} \\ &\quad +\boldsymbol{\phi}_X(\boldsymbol{Y})_{il}\boldsymbol{\phi}_X(\boldsymbol{Y})_{jl}\boldsymbol{R}(t)_{ll}\} \\ &= -2\boldsymbol{R}_{ll}(t)\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_{jl}\boldsymbol{\phi}_X(\boldsymbol{Y})_{il}\} \end{aligned} \quad (17)$$

where

$$\boldsymbol{\phi}_X(\boldsymbol{y}) = \mathbb{E}\{(\boldsymbol{X} - \mathbb{E}\{\boldsymbol{X}|\boldsymbol{y}\})(\boldsymbol{X} - \mathbb{E}\{\boldsymbol{X}|\boldsymbol{y}\})^T|\boldsymbol{y}\}. \quad (18)$$

Note that $\boldsymbol{\phi}_X(\boldsymbol{y})$ depends on t through $\boldsymbol{Y}(t) = \boldsymbol{R}(t)\boldsymbol{X} + \boldsymbol{N}$. The second equality in equation (17) is due to the fact that $\boldsymbol{R}(t)$ is diagonal. Thus, we can write the derivative of $\boldsymbol{E}_{ij}(t)$ as

$$\begin{aligned} D_t\boldsymbol{E}_{ij}(t) &= -2\sum_l \boldsymbol{R}_{ll}(t)\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_{jl}\boldsymbol{\phi}_X(\boldsymbol{Y})_{il}\}\boldsymbol{R}'_{ll}(t) \\ &= -2\sum_l \boldsymbol{R}_{ll}(t)\boldsymbol{R}'_{ll}(t)\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_{jl}\boldsymbol{\phi}_X(\boldsymbol{Y})_{il}\} \\ &= -2\sum_l \boldsymbol{B}_{ll}(t)\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_{jl}\boldsymbol{\phi}_X(\boldsymbol{Y})_{il}\} \end{aligned} \quad (19)$$

recalling that $\boldsymbol{B}_{ll}(t) = \boldsymbol{R}_{ll}(t)\boldsymbol{R}'_{ll}(t)$. We can put this expression into a matrix form as follows:

$$D_t\boldsymbol{E}(t) = -2\sum_l \boldsymbol{B}_{ll}(t)\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_l\boldsymbol{\phi}_X(\boldsymbol{Y})_l^T\} \quad (20)$$

where $\boldsymbol{\phi}_X(\boldsymbol{Y})_l$ is the l^{th} column of the matrix $\boldsymbol{\phi}_X(\boldsymbol{Y})$. Using the fact that for a Gaussian input distribution $\boldsymbol{\phi}_X(\boldsymbol{Y})$ does not depend on $\boldsymbol{Y}$ and, thus, $\boldsymbol{\phi}_X(\boldsymbol{Y}) = \mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})\} = \boldsymbol{E}^G(t)$ [14], we can obtain the following lower bound on the derivative of the matrix $\boldsymbol{Q}(t)$:

$$\begin{aligned} &\boldsymbol{Q}'(t) \\ &= 2\sum_l \boldsymbol{B}_{ll}(t)\left(\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_l\boldsymbol{\phi}_X(\boldsymbol{Y})_l^T\} - \boldsymbol{E}_l^G(\boldsymbol{E}_l^G)^T\right) \\ &\succeq 2\sum_l \boldsymbol{B}_{ll}(t)\left(\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_l\}\mathbb{E}\{\boldsymbol{\phi}_X(\boldsymbol{Y})_l\}^T - \boldsymbol{E}_l^G(\boldsymbol{E}_l^G)^T\right) \\ &= 2\sum_l \boldsymbol{B}_{ll}(t)\left(\boldsymbol{E}_l\boldsymbol{E}_l^T - \boldsymbol{E}_l^G(\boldsymbol{E}_l^G)^T\right) \\ &= 2\left(\boldsymbol{E}(t)\boldsymbol{B}(t)\boldsymbol{E}^T(t) - \boldsymbol{E}_G(t)\boldsymbol{B}(t)\boldsymbol{E}_G^T(t)\right) \end{aligned}$$

where the inequality is due to Jensen. ■

Let us fix $t_0 \geq 0$ and consider the generalized eigenvalue decomposition [15] on $(\boldsymbol{E}_G(t_0), \boldsymbol{E}(t_0))$. [1] Thus, there exists an invertible matrix $\boldsymbol{V}_0$ such that,

$$\begin{aligned} \boldsymbol{E}_G(t_0) &= \boldsymbol{V}_0^T\boldsymbol{V}_0 \\ \boldsymbol{E}(t_0) &= \boldsymbol{V}_0^T\boldsymbol{\Sigma}_0\boldsymbol{V}_0 \end{aligned} \quad (21)$$

where $\boldsymbol{\Sigma}_0$ is a positive semi-definite diagonal matrix. Thus,

$$\boldsymbol{Q}(t_0) = \boldsymbol{E}_G(t_0) - \boldsymbol{E}(t_0) = \boldsymbol{V}_0^T\left(\boldsymbol{I} - \boldsymbol{\Sigma}_0\right)\boldsymbol{V}_0 \quad (22)$$

and the following matrix:

$$\widetilde{\boldsymbol{Q}}(t_0) = \boldsymbol{V}_0^{-T}\boldsymbol{Q}(t_0)\boldsymbol{V}_0^{-1} \quad (23)$$

is diagonal. By defining $\boldsymbol{C}_0 = \boldsymbol{V}_0\boldsymbol{B}(t_0)\boldsymbol{V}_0^T$, we can rewrite the lower bound attained in Lemma 3 as follows:

$$\boldsymbol{Q}'(t) \succeq 2\boldsymbol{V}_0^T\left(\boldsymbol{\Sigma}_0\boldsymbol{C}_0\boldsymbol{\Sigma}_0 - \boldsymbol{C}_0\right)\boldsymbol{V}_0. \quad (24)$$

Our main result is the following:

Theorem 1. *Each eigenvalue of $\boldsymbol{Q}(t)$ crosses the horizontal axis at most once.*

Proof: Consider the new matrix function $\widetilde{\boldsymbol{Q}}(t) = \boldsymbol{V}_0^{-T}\boldsymbol{Q}(t)\boldsymbol{V}_0^{-1}$, which, from Sylvester's law of inertia, has the same number of positive, negative, and zero eigenvalues as $\boldsymbol{Q}(t)$.

Note that $\widetilde{\boldsymbol{Q}}(t_0) = \boldsymbol{I} - \boldsymbol{\Sigma}_0$ is a diagonal matrix and, thus, we can apply Corollary 2 to obtain a lower bound on the derivative of the eigenvalues of $\widetilde{\boldsymbol{Q}}(t)$ evaluated at t_0:

$$\begin{aligned} \left.\frac{\mathrm{d}\lambda_i\left(\widetilde{\boldsymbol{Q}}(t)\right)}{\mathrm{d}t}\right|_{t=t_0} &= \left.\frac{\mathrm{d}\left[\widetilde{\boldsymbol{Q}}(t)\right]_{ii}}{\mathrm{d}t}\right|_{t=t_0} = \left[\boldsymbol{V}_0^{-T}\boldsymbol{Q}'(t_0)\boldsymbol{V}_0^{-1}\right]_{ii} \\ &\geq 2([\boldsymbol{\Sigma}_0]_{ii}[\boldsymbol{C}_0]_{ii}[\boldsymbol{\Sigma}_0]_{ii} - [\boldsymbol{C}_0]_{ii}) \end{aligned} \quad (25)$$

where in (25) we applied the lower bound given in Lemma 3.

Now, let us particularize the bound obtained in (25) to the non-positive eigenvalues of $\widetilde{\boldsymbol{Q}}(t_0)$, i.e., those that fulfill that

[1] In the generalized eigenvalue decomposition we have considered that $\boldsymbol{E}_G \succ \boldsymbol{0}$. A sufficient condition for $\boldsymbol{E}_G \succ \boldsymbol{0}$ is that the covariance of the Gaussian input distribution is non-singular and that $\boldsymbol{R}(t)$ is non-singular.

$\lambda_i\big(\widetilde{\boldsymbol{Q}}(t)\big) \leq 0$ which implies that $[\boldsymbol{\Sigma}_0]_{ii} \geq 1$, from which it follows that

$$\left.\frac{\mathrm{d}\lambda_i\big(\widetilde{\mathbf{Q}}(t)\big)}{\mathrm{d}t}\right|_{t=t_0} \geq 2([\boldsymbol{\Sigma}_0]_{ii}[\boldsymbol{C}_0]_{ii}[\boldsymbol{\Sigma}_0]_{ii} - [\boldsymbol{C}_0]_{ii}) \geq 0 \quad (26)$$

where we have used the fact that $\boldsymbol{B}(t) \succeq \mathbf{0}$.

The last result implies that, in a sufficiently small neighborhood of t_0, the non-positive eigenvalues of $\widetilde{\boldsymbol{Q}}(t)$ are non-decreasing functions of t. Consequently, from the continuity of the eigenvalues, the number of negative eigenvalues of $\widetilde{\boldsymbol{Q}}(t)$ cannot increase.

Now, taking into account that the number of positive, zero, and negative eigenvalues is preserved under the transformation $\widetilde{\boldsymbol{Q}}(t) \mapsto \boldsymbol{Q}(t)$ we will informally show that a zero eigenvalue of $\boldsymbol{Q}(t)$ cannot become negative (a complete formal proof is given in [16]). This will prove that each eigenvalue of $\boldsymbol{Q}(t)$ crosses the horizontal axis at most once. We will prove this result by contradiction. Let's assume that there is a zero eigenvalue of $\boldsymbol{Q}(t_0)$ that becomes negative for $t > t_0$. Since the number of negative eigenvalues cannot increase, there must be at least one negative eigenvalue at t_0 that increases to zero. If we examine the sign of the eigenvalues at $t_0 + \Delta$ for a sufficiently small Δ, we know that the zero eigenvalue has to be negative, however the negative eigenvalue (for sufficiently small Δ) is also still negative. Thus, we will have an increase in the number of negative eigenvalues, contradicting the property of no increase in negative eigenvalues of $\widetilde{\boldsymbol{Q}}(t)$. This shows that a zero eigenvalue cannot become negative in $\boldsymbol{Q}(t_0)$. ■

The following corollary is a simple consequence from Theorem 1.

Corollary 1. *If, for a given t', the function $\boldsymbol{Q}(t')$ fulfils that $\boldsymbol{Q}(t') \succeq \mathbf{0}$ then for all $t \geq t'$ we also have that $\boldsymbol{Q}(t) \succeq \mathbf{0}$.*

Note that, by restricting the input distributions to be *i.i.d.*, the matrix $\boldsymbol{Q}(t)$ is a diagonal matrix for all t. Thus, the single crossing property of the eigenvalues simplifies to a single crossing property of the diagonal values, as expected due to the scalar single crossing property [9], [10]. However, in the general case, where the input distribution is arbitrary, the multivariate unique crossing property does not follow directly from the scalar case.

V. Connecting to the Mutual Information

As in the scalar scenario, our goal is to use the "single crossings" in order to derive results regarding the mutual information. According to the I-MMSE relationship (5), we would like to examine the following function:

$$\mathrm{tr}\{\boldsymbol{B}(t)\boldsymbol{E}_G(t)\} - \mathrm{tr}\{\boldsymbol{B}(t)\boldsymbol{E}(t)\} = \mathrm{tr}\{\boldsymbol{B}(t)\boldsymbol{Q}(t)\}. \quad (27)$$

Since the trace is the sum of the eigenvalues, we need the following lemma, that extends our results regarding the eigenvalues of $\boldsymbol{Q}(t)$ to the eigenvalues of $\boldsymbol{B}(t)\boldsymbol{Q}(t)$ for a positive semi-definite diagonal matrix $\boldsymbol{B}(t)$.

Lemma 4. *Each eigenvalue of $\boldsymbol{B}(t)\boldsymbol{Q}(t)$ crosses the horizontal axis at most once.*

Proof: Using,

$$\lambda_i\{\boldsymbol{B}(t)\boldsymbol{Q}(t)\} = \lambda_i\{\boldsymbol{B}^{\frac{1}{2}}(t)\boldsymbol{Q}(t)\boldsymbol{B}^{\frac{1}{2}}(t)\} \quad (28)$$

with the fact that $\boldsymbol{B}(t)$ is diagonal and positive semi-definite, we again have the eigenvalues of a congruent transformation. Thus, the extension of the previous claim follows directly. ■

In order to use our results regarding the function $\boldsymbol{Q}(t)$ we need the following lemma:

Lemma 5. *For any t', there exists a Gaussian input covariance matrix $\boldsymbol{C}_G$ such that the following holds:*

1) $\boldsymbol{C}_G \preceq \boldsymbol{C}_X$
2) $I(\boldsymbol{X};\boldsymbol{Y}(t')) = I(\boldsymbol{X}_G;\boldsymbol{Y}_G(t'))$
3) $\boldsymbol{Q}(t') \succeq \mathbf{0}$

where $\boldsymbol{Y}(t') = \boldsymbol{R}(t')\boldsymbol{X} + \boldsymbol{N}$ and $\boldsymbol{Y}_G(t') = \boldsymbol{R}(t')\boldsymbol{X}_G + \boldsymbol{N}$.

Proof: Due to the space limitations we will give only a sketch of the proof. For full details see [16]. From the third requirement we have:

$$\boldsymbol{Q}(t') = \boldsymbol{E}_G(t') - \boldsymbol{E}(t') \equiv \boldsymbol{J} \succeq \mathbf{0}. \quad (29)$$

Furthermore, we can define:

$$\boldsymbol{C} \equiv \boldsymbol{E}_L(t') - \boldsymbol{E}(t') \succeq \mathbf{0} \quad (30)$$

where $\boldsymbol{E}_L(t')$ is the error covariance matrix assuming an optimal linear estimator. When $\boldsymbol{J} = \mathbf{0}$ ($\boldsymbol{E}_G(t') = \boldsymbol{E}(t')$) we have that $\boldsymbol{Q}(t') = \mathbf{0}$. According to Theorem 1 we have that all eigenvalues are non-positive for $t \leq t'$. Furthermore, due to Lemma 4 we conclude that the eigenvalues of $\boldsymbol{B}(t)\boldsymbol{Q}(t)$ are also non-positive for all $t \leq t'$ and we can conclude, using (5), that $I(\boldsymbol{X}_G;\boldsymbol{Y}_G(t')) \leq I(\boldsymbol{X};\boldsymbol{Y}(t'))$. If $\boldsymbol{J} = \boldsymbol{C}$ we have that $\boldsymbol{C}_G = \boldsymbol{C}_X$ in which case we have $I(\boldsymbol{X}_G;\boldsymbol{Y}_G(t')) \geq I(\boldsymbol{X};\boldsymbol{Y}(t'))$. In order to comply with the first requirement we need to require that $\boldsymbol{J} \preceq \boldsymbol{C}$. Thus, requirements 1 and 3 can be written using $\boldsymbol{J}$ and $\boldsymbol{C}$, defined in equations (29) and (30) respectively, and we have the following:

$$I(\boldsymbol{X}_G;\boldsymbol{Y}_G(t'))\Big|_{\boldsymbol{J}=\mathbf{0}} \leq I(\boldsymbol{X};\boldsymbol{Y}(t')) \leq I(\boldsymbol{X}_G;\boldsymbol{Y}_G(t'))\Big|_{\boldsymbol{J}=\boldsymbol{C}}.$$

The question is whether there exists such a $\boldsymbol{J}$ that will also attain $I(\boldsymbol{X}_G;\boldsymbol{Y}_G(t')) = I(\boldsymbol{X};\boldsymbol{Y}(t')) \equiv \alpha$. Both upper and lower bound can be expressed using the function $r(t) = \frac{1}{2}\log\frac{|\boldsymbol{A}|}{|\boldsymbol{B}+\boldsymbol{\Delta}\nu|}$ which is continuous and monotonically decreasing in ν for $0 \leq \nu \leq 1$ for $\boldsymbol{A} \succ \mathbf{0}$, $\boldsymbol{B} \succ \mathbf{0}$ and $\boldsymbol{\Delta} \succeq \mathbf{0}$ [17]. Thus, there exists a $0 \leq \nu^* \leq 1$ that attains equality. That is, $\boldsymbol{J}^* = (1-\nu^*)\boldsymbol{C}$ attains all three requirement. ■

VI. Application: parallel degraded MIMO Gaussian BC

As an example for the usage of these results we examine the parallel *degraded* Gaussian BC. We first note that the results attained so far have also been extended to the conditioned case where $(\boldsymbol{X}, U)$ are jointly distributed and $U - \boldsymbol{X} - \boldsymbol{Y}$ forms a Markov chain, but are omitted here due to space limitations.

The conditioned MMSE is defined as:

$$\begin{aligned}\boldsymbol{E}^u(t) &= \mathbb{E}\{(\boldsymbol{X}-\mathbb{E}\{\boldsymbol{X}|\boldsymbol{R}(t)\boldsymbol{X}+\boldsymbol{N},U\}) \\ &\quad (\boldsymbol{X}-\mathbb{E}\{\boldsymbol{X}|\boldsymbol{R}(t)\boldsymbol{X}+\boldsymbol{N},U\})^T\} \quad (31)\end{aligned}$$

and the conditioned matrix $\boldsymbol{Q}^u(t) = \boldsymbol{E}^G(t) - \boldsymbol{E}^u(t)$, which is the difference between the MMSE matrix assuming a general Gaussian input distribution independent of U, and the conditioned MMSE matrix.

We consider the *degraded* parallel Gaussian BC channel:

$$\begin{aligned}\boldsymbol{Y}_1[m] &= \boldsymbol{H}_1\boldsymbol{X}[m]+\boldsymbol{N}_1[m] \\ \boldsymbol{Y}_2[m] &= \boldsymbol{H}_2\boldsymbol{X}[m]+\boldsymbol{N}_2[m] \quad (32)\end{aligned}$$

where $\boldsymbol{N}_1[m]$ and $\boldsymbol{N}_2[m]$ are standard additive Gaussian noise vectors, $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are diagonal positive definite matrices such that $\boldsymbol{H}_1 \preceq \boldsymbol{H}_2$. The channel input satisfies the covariance constraint: $\mathbb{E}\{\boldsymbol{X}\boldsymbol{X}^T\} \preceq \boldsymbol{S}$, where $\boldsymbol{S}$ is some positive definite matrix.

One way of proving that the Gaussian input achieves the capacity region is by using the single-letter expression [18]:

$$\begin{aligned}R_1 &\leq I(U;\boldsymbol{Y}_1) \\ R_2 &\leq I(\boldsymbol{X};\boldsymbol{Y}_2|U) \quad (33)\end{aligned}$$

where U is an auxiliary random variable over a certain alphabet that satisfies the Markov relation $U-\boldsymbol{X}-(\boldsymbol{Y}_1,\boldsymbol{Y}_2)$. This was done for the scalar Gaussian BC in [9], [10]. We will try to follow similar steps for the *degraded* parallel Gaussian channel.

Assume a pair $(\boldsymbol{X},U)$ with covariance matrix $\boldsymbol{C}_X$. Using the conditioned version of Lemma 5 we know that there exists a Gaussian distribution, with covariance $\boldsymbol{B} \preceq \boldsymbol{S}$, with the following properties:

$$I(\boldsymbol{X};\boldsymbol{H}_1\boldsymbol{X}+\boldsymbol{N}|U) = I(\boldsymbol{X}_G;\boldsymbol{H}_1\boldsymbol{X}_G+\boldsymbol{N}) \quad (34)$$

$$\begin{aligned}&I(\boldsymbol{X}_G;\boldsymbol{H}_2\boldsymbol{X}_G+\boldsymbol{N}) - I(\boldsymbol{X};\boldsymbol{H}_2\boldsymbol{X}+\boldsymbol{N}|U) \\ &= \int_{t=0}^{t_2} \mathrm{tr}\{\boldsymbol{B}(t)\boldsymbol{Q}^u(t)\}dt \\ &= \int_{t=0}^{t_1} \mathrm{tr}\{\boldsymbol{B}(t)\boldsymbol{Q}^u(t)\}dt + \int_{t_1}^{t_2} \mathrm{tr}\{\boldsymbol{B}(t)\boldsymbol{Q}^u(t)\}dt \quad (35) \\ &= 0 + \int_{t_1}^{t_2} \mathrm{tr}\{\boldsymbol{B}(t)\boldsymbol{Q}^u(t)\}dt \geq 0 \quad (36)\end{aligned}$$

where (35) is due to (34), and the inequality is due to the fact that $\boldsymbol{Q}^u(t) \succeq \boldsymbol{0}$ for all $t \geq t_1$ and Lemma 4. Thus, we have a Gaussian distribution that complies with a covariance constraint and also,

$$\begin{aligned}&I(\boldsymbol{X};\boldsymbol{H}_1\boldsymbol{X}+\boldsymbol{N}|U) = I(\boldsymbol{X}_G;\boldsymbol{H}_1\boldsymbol{X}_G+\boldsymbol{N}) \\ &I(\boldsymbol{X};\boldsymbol{H}_2\boldsymbol{X}+\boldsymbol{N}|U) \leq I(\boldsymbol{X}_G;\boldsymbol{H}_2\boldsymbol{X}_G+\boldsymbol{N}) \quad (37)\end{aligned}$$

assuming a parallel *degraded* model, that is, $\boldsymbol{0} \prec \boldsymbol{H}_1 \preceq \boldsymbol{H}_2$. Now, substituting the above into the region given in equation (33) we obtain the following region:

$$\begin{aligned}R_1 &\leq I(U;\boldsymbol{Y}_1) = I(\boldsymbol{X};\boldsymbol{Y}_1) - I(\boldsymbol{X};\boldsymbol{Y}_1|U) \\ &\leq \frac{1}{2}\log|\boldsymbol{I}+\boldsymbol{H}_1\boldsymbol{S}\boldsymbol{H}_1^T| - \frac{1}{2}\log|\boldsymbol{I}+\boldsymbol{H}_1\boldsymbol{B}\boldsymbol{H}_1^T| \\ &= \frac{1}{2}\log\frac{|\boldsymbol{I}+\boldsymbol{H}_1\boldsymbol{S}\boldsymbol{H}_1^T|}{|\boldsymbol{I}+\boldsymbol{H}_1\boldsymbol{B}\boldsymbol{H}_1^T|} \quad (38)\end{aligned}$$

$$R_2 \leq I(\boldsymbol{X};\boldsymbol{Y}_2|U) \leq \frac{1}{2}\log|\boldsymbol{I}+\boldsymbol{H}_2\boldsymbol{B}\boldsymbol{H}_2^T|. \quad (39)$$

This concludes the converse part of the proof. The achievability is well-known using Gaussian superposition coding.

REFERENCES

[1] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[2] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 141–154, January 2006.

[3] D. Guo, S. Shamai (Shitz), and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE International Symposium on Information Theory (ISIT 2006)*, Seattle, WA, July 9-14 2006.

[4] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.

[5] F. Pérez-Cruz, M. R. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1070–1084, March 2010.

[6] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, December 2008.

[7] M. Peleg, A. Sanderovich, and S. Shamai (Shitz), "On extrinsic information of good codes operating over Gaussian channels," *European Transactions on Telecommunications*, vol. 18, no. 2, pp. 133–139, 2007.

[8] A. M. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4295–4297, September 2006.

[9] D. Guo, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," in *Proc. IEEE International Symposium on Information Theory (ISIT 2008)*, Toronto, ON, Canada, July 6-11 2008.

[10] D. Guo, Y. Wu, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," submitted to the *IEEE Transactions on Information Theory*, 2010.

[11] E. Ekrem and S. Ulukus, "Secrecy capacity region of the Gaussian multi-receive wiretap channel," in *Proc. IEEE International Symposium on Information Theory (ISIT 2009)*, Seoul, Korea, June 28 - July 3 2009.

[12] R. Bustin, R. Liu, H. V. Poor, and S. Shamai (Shitz), "An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel," *EURASIP Journal on Wireless Communications and Networking, Special Issue on Wireless Physical Layer Security*, August 2009, available online: http://www.hindawi.com/journals/wcn/2009/370970.html.

[13] J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics, 3^{rd} edition.* New York, Wiley, 2007.

[14] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3613–3628, August 2009.

[15] G. Strang, *Linear Algebra and Its Applications*. Wellesley, MA: Wellesley-Cambridge Press, 1998.

[16] R. Bustin, M. Payaró, D. P. Palomar, and S. Shamai (Shitz), "On MMSE properties and I-MMSE implications in parallel MIMO Gaussian channels," in preparation.

[17] E. Ekrem and S. Ulukus, "The secrecy capacity region of the Gaussian MIMO multi-receive wiretap channel," submitted to the *IEEE Transactions on Information Theory*, March 2009, available at: arXiv:0903.3096.

[18] T. M. Cover, "Comments on broadcast channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2524–2530, October 1998.